Distinct effects on mRNA export factor GANP underlie neurological disease phenotypes and alter gene expression depending on intron content

Woldegebriel, Rosa

2020-05-01

Woldegebriel , R , Kvist , J , Andersson , N , Ounap , K , Reinson , K , Wojcik , M H , Bijlsma , E K , Hoffer , M J , Ryan , M M , Stark , Z , Walsh , M , Cuppen , I , van den Boogaard , M-J H , Bharucha-Goebel , D , Donkervoort , S , Winchester , S , Zori , R , Bonnemann , C G , Maroofian , R , O'Connor , E , Houlden , H , Zhao , F , Carpen , O , White , M , Sreedharan , J , Stewart , M , Ylikallio , E & Tyynismaa , H 2020 , ' Distinct effects on mRNA export factor GANP underlie neurological disease phenotypes and alter gene expression depending on intron content ' , Human Molecular Genetics , vol. 29 , no. 9 , pp. 1426-1439 . https://doi.org/10.1093/hmg/ddaa051

http://hdl.handle.net/10138/318305
https://doi.org/10.1093/hmg/ddaa051

Human Molecular Genetics, 2020, Vol. 29, No. 9 1426–1439

doi: 10.1093/hmg/ddaa051
Advance Access Publication Date: 23 March 2020
General Article

GENERAL ARTICLE

Distinct effects on mRNA export factor GANP underlie neurological disease phenotypes and alter gene expression depending on intron content

Rosa Woldegebriel[1,2], Jouni Kvist[1], Noora Andersson[3,20], Katrin Õunap[4,5], Karit Reinson[4,5], Monica H. Wojcik[6,7], Emilia K. Bijlsma[8], Mariëtte J. V. Hoffer[8], Monique M. Ryan[9,10,11], Zornitza Stark[9,11], Maie Walsh[9], Inge Cuppen[12], Marie-José H. van den Boogaard[13], Diana Bharucha-Goebel[14,15], Sandra Donkervoort[14], Sara Winchester[16], Roberto Zori[17], Carsten G. Bönnemann[14], Reza Maroofian[18], Emer O'Connor[18], Henry Houlden[18], Fang Zhao[19], Olli Carpén[3,20], Matthew White[2], Jemeen Sreedharan[2], Murray Stewart[21], Emil Ylikallio[1,22,*] and Henna Tyynismaa[1,23,24,*]

[1]Stem Cells and Metabolism Research Program, Research Programs Unit, University of Helsinki, 00290 Helsinki, Finland, [2] Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, [3]Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland, [4]Department of Clinical Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia, [5]Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia, [6] Broad Institute of MIT and Harvard, Cambridge, MA, USA, [7]Divisions of Genetics and Genomics and Newborn Medicine, Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA, [8]Department of Clinical Genetics, Leiden University Medical Centre, Leiden, the Netherlands, [9] Murdoch Children's Research Institute, Melbourne 3052, Australia, [10] Royal Children's Hospital, Melbourne 3052, Australia, [11]Department of Paediatrics, The University of Melbourne, Melbourne 3052, Australia, [12]Department of Pediatric Neurology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands, [13]Department of Genetics, University Medical Center Utrecht, The Netherlands, [14]Neuromuscular and Neurogenetic Disorders of Childhood Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA, [15] Division of Neurology, Children's National Health System, Washington, DC, USA, [16] Child Neurology Center of Northwest Florida, Pensacola, FL, USA, [17]Division of Genetics and Metabolism, University of Florida, Gainesville, FL, USA, [18]Department of Neuromuscular Disorders, UCL Institute of Neurology, London WC1N 3BG, UK, [19]Department of Pathology and Genetics, HUSLAB Laboratories, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, [20]Research Program in Systems Oncology, University of Helsinki, Helsinki, Finland, [21]MRC Laboratory of Molecular Biology, Francis Crick Ave, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK, [22]Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, 00290 Helsinki, Finland, [23]Department of Medical and Clinical Genetics, University of Helsinki, 00290 Helsinki, Finland and [24]Neuroscience Center, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland

*To whom correspondence should be addressed at: Biomedicum Helsinki, r.C525b, Haartmaninkatu 8, 00290 Helsinki, Finland. Tel: +358 2941 25654; Email: henna.tyynismaa@helsinki.fi (Henna Tyynismaa); emil.ylikallio@helsinki.fi (Emil Ylikallio)

Received: January 29, 2020. **Revised:** March 17, 2020. **Accepted:** March 19, 2020

Abstract

Defects in the mRNA export scaffold protein GANP, encoded by the *MCM3AP* gene, cause autosomal recessive early-onset peripheral neuropathy with or without intellectual disability. We extend here the phenotypic range associated with *MCM3AP* variants, by describing a severely hypotonic child and a sibling pair with a progressive encephalopathic syndrome. In addition, our analysis of skin fibroblasts from affected individuals from seven unrelated families indicates that disease variants result in depletion of GANP except when they alter critical residues in the Sac3 mRNA binding domain. GANP depletion was associated with more severe phenotypes compared with the Sac3 variants. Patient fibroblasts showed transcriptome alterations that suggested intron content-dependent regulation of gene expression. For example, all differentially expressed intronless genes were downregulated, including *ATXN7L3B*, which couples mRNA export to transcription activation by association with the TREX-2 and SAGA complexes. Our results provide insight into the molecular basis behind genotype-phenotype correlations in *MCM3AP*-associated disease and suggest mechanisms by which GANP defects might alter RNA metabolism.

Introduction

We and others have recently established that biallelic variants in minichromosome maintenance complex component 3-associated protein (*MCM3AP*) cause peripheral neuropathy with or without impaired intellectual development (PNRIID, MIM 618124) (1,2,3,4). This childhood-onset neurologic syndrome is progressive and causes distal motor impairment and gait difficulties, often resulting in wheelchair dependency and in difficulties using proximal limbs. Most affected individuals have impaired intellectual development or learning difficulties, although some have normal cognition. Electrophysiological testing and sural nerve biopsy are most compatible with an axonal sensorimotor neuropathy, whereas some patients have signs of demyelination. Additional features may include upper motor neuron signs, cerebellar ataxia, eye movement abnormalities, characteristic claw hands and foot deformities and scoliosis. Moreover, the phenotypic range of *MCM3AP*-associated disease was expanded by a report of patients who developed peripheral neuropathy and multiple sclerosis-type disease (5) and segregation of a heterozygous variant in a family with periodic paralysis (6).

MCM3AP encodes for germinal center-associated nuclear protein (GANP), which is a 218 kDa ubiquitously expressed protein (7,8). GANP forms a scaffold for several interacting proteins in the Transcription Export-2 (TREX-2) complex that facilitates the nuclear export of a subset of mRNAs through nuclear pore complexes (NPCs) (7,8,9,10). NPCs on the nuclear envelope are composed of nucleoporins, which form channels that have phenylalanine-glycine (FG) repeats in the middle, forming a mesh-like network that mediates selective exchange of components between the nucleus and cytoplasm. GANP is a multi-domain protein in which the N-terminus contains a cluster of six nucleoporin-like FG repeats. In the middle, GANP has a Sac3 (Suppressor of actin 3) mRNA binding domain that shows a high homology from yeast to metazoans (11,12,13). In yeast, Sac3 interacts physically with nucleoporins to mediate recruitment of the mRNA export machinery to the NPCs (11,12,14). At its C-terminus, GANP has an acetyltransferase domain that is crucial for localization to the nuclear envelope and has been described as involved in acetylation of MCM3 during DNA replication (15,16).

The export of mRNAs is tightly coupled to their transcription and processing, with strict quality control preventing the export of misprocessed messengers, in particular pre-mRNAs with unspliced introns (17). Thus, a defect in a particular step from transcription, mRNA maturation and termination to export is likely to have direct consequences on both upstream and downstream steps. In human cells, two large transcription export complexes TREX and TREX-2 link transcription to mRNA export. They recognize mRNAs during early stages of biogenesis and transfer the mature, spliced mRNAs to the NXF1:NXT1 nuclear export factor, which facilitates their export to the cytoplasm through interactions with the FG repeats of nucleoporins. TREX is recruited to mRNAs during splicing (18). TREX-2 is based on a GANP scaffold to which PCID2, DSSI, ENY2 and CETN2/3 bind (10). TREX-2 is also associated with transcription through the interaction of ENY2 and GANP with RNA polymerase II (10). The multiprotein complexes involved in transcription, processing and export share protein components and modules reciprocally, allowing coordination or modulation of these processes (19). For example, ENY2 in the TREX-2 complex is also utilized by the histone deubiquitinase (UBA) module of the large transcriptional co-activator SAGA (Spt-Ada-Gcn5 acetyltransferase) complex (20,21). In *Drosophila melanogaster*, TREX-2 and SAGA are proposed to co-regulate transcription and mRNA export (22).

Subsets of mRNAs can be exported selectively, but the molecular machinery involved and their regulation are unclear. In a fly study, NXF1 was found to export most but not all mRNAs (23). On the contrary, a deficiency in THOC5, a component of the TREX complex, decreased the cytoplasmic expression of only a small subset of mRNA enriched for those required for hematopoietic development (24), whereas TREX-2, through GANP, promoted the nuclear export of specific mRNA classes required for gene expression itself, such as RNA processing, splicing, mRNP and ribosome biogenesis (25). Furthermore, transcripts selectively regulated by GANP in colon cancer cells had fewer introns than those regulated by GANP and NXF1 together or NXF1 only, suggesting that GANP may facilitate rapid changes in gene expression by regulating the export of genes that require less processing (9).

We have previously suggested a genotype–phenotype correlation for pathogenic variants in *MCM3AP* whereby patients with missense variants in Sac3 domain present with milder motor phenotypes compared with patients who have compound heterozygous variants affecting other parts of the protein (26). To investigate the molecular consequences of the different mutation types on GANP localization and function, we have studied here skin fibroblasts from multiple patients with *MCM3AP*-associated disease. We found that more severe

phenotypes were characterized by near complete loss of GANP at the nuclear envelope, whereas missense variants in the Sac3 domain of GANP did not lead to GANP depletion. Furthermore, GANP defects impacted the overall mRNA expression in patient fibroblasts, with differentially expressed genes showing intron content dependency. Overall, the present study suggests that GANP deficiency impairs transcription and/or export of specific mRNAs, which provides novel insight into the roles of GANP in human disease.

Results

Novel *MCM3AP* variants and the associated phenotypes

We identified two new families, one in the Netherlands and one in Estonia, with variants in *MCM3AP*, by genome or exome sequencing. Clinical features of these and previously reported patients with autosomal recessive disease are listed in Supplementary material Table S1.

The affected individual in Dutch family (NL2) is a girl with severe hypotonia and intellectual disability. The patient is now 4 years 9 months old, born as the first child of healthy, non-consanguineous parents (father Dutch, mother Thai). From the age of 3 months, there were concerns about motor development, as she had poor head control. She was referred to a pediatric neurologist at the age of 7 months. Neurologic examination showed no signs of motor and sensory neuropathy, and her neurological features rather suggested a central defect. At the age of 16 months, she was referred to a clinical geneticist because of motor delay, hypotonia, swallowing problems and abnormal (dyskinetic) movements. She was not dysmorphic.

At the age of 2 years 2 months, she had slow motor progression and spoke some words. Her best motor milestone was rolling at the age of 2 years. Attempts to crawl were hampered by a cast for severe scoliosis that she had developed. At the age of 3 years 5 months, she spoke only six words. At the age of 4 years 8 months—after a period of viral infections—she started to regress. She could no longer move her legs against gravity. She became short of breath when lying on her back. She had recurrent upper airway infections and progressive severe scoliosis. At the age of 4 years, she required placement of a gastrostomy tube because of poor weight gain and vomiting. During a period of high fever, her respiratory capacity dropped and she was placed on ventilatory support, which is still required at age 4 years 9 months.

A brain MRI at the age of 10 months was normal. At the age of 4 years 9 months, hyperintensity of the basal ganglia and mild ventriculomegaly were seen. The radiologic finding was suggestive of a metabolic abnormality.

An EMG at the age of 4 years 9 months showed affected motor and sensory fibers, with severe polyneuropathy.

Array analysis showed three copy number variants, paternal duplication of 4p and deletion of 5q, and maternal deletion of 7q. Whole exome analysis identified a maternally inherited intronic variant c.4290 + 3G > T in *MCM3AP*, predicted to affect the splicing of exon 20, and a paternally inherited frameshift variant c.3476_3477del (p.Gln1159Argfs*8), predicted to lead to a premature termination (Fig. 1A). Neither variant is found in the ExAC variant database.

The Estonian family (EST) has two affected siblings. Both are in their late twenties (brother, EST1; sister, EST2). They were born at full term to non-consanguineous Estonian parents following an unremarkable pregnancy. Their birth weights were relatively small (−1.75SD), but the length was 0SD and the head circumferences were not known. Mild psychomotor developmental delay was noticed in the first year of life (head control at age 5 months; standing independently at the age of 12 months). Both siblings had severe generalized epilepsy and mild spastic diplegia, diagnosed at the age of 3.5 years (EST1) and 8 years (EST2). They both began to learn at a conventional school with a conventional learning program. However, already at the age of 11 years (EST2) and 12 years (EST1), progressive encephalopathy was diagnosed in addition to polyneuropathy, and soon drug resistant epilepsy presented. The EEG showed slow background activity accompanied by frequent epileptiform activity. At this age, the head circumference, weight and length were normal. Over the years, the symptoms constantly progressed: most of the learned skills disappeared; both patients developed muscle atrophy, spasticity and joint contraction; ataxia and dysarthria; aggression and behavioral problems (mostly EST1). At the age of 26–27 years, both individuals are mostly bedridden. They can sit but not stand or walk. They have moderate to severe intellectual disability and drug-resistant epilepsy (1–10 seizures per day).

The EST siblings have the heterozygous c.3814G > A (p.Val1272Met) missense variant, which we previously reported in a Finnish family (2), and a maternally inherited novel frameshift variant c.5423delG (p.Gly1808Aspfs*5) (Fig. 1A). Comparison of the sequencing data of the previously reported Finnish patients and the Estonian patients suggested that affected individuals from these families have the same haplotype on the c.3814G > A (p.Val1272Met) allele as each patient also shared two rare intronic variants in the gene (hg19: 21-47 662 706 T > C and 21-47690191G > A). These variants as well as the disease variant (21-47676821C > T) are enriched in the Finnish population, each having an allele frequency of about 0.0025 according to gnomAD database.

We previously found a decrease in mRNA levels of *MCM3AP* in patient fibroblasts (2). We now assessed *MCM3AP* mRNA levels in the skin fibroblasts of the NL2, EST1 and EST2 patients by quantitative PCR (qPCR), normalized to Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). All three patients had approximately 30% of *MCM3AP* mRNA remaining in comparison with the expression level in healthy control fibroblasts (Fig. 1B). As all patients carried frameshift variants on one allele, which are expected to result in mRNA degradation, the qPCR result suggested that the second variant in each case also affected mRNA stability.

GANP missense variants in Sac3 domain are clustered and probably impair interaction with RNA

Disease associated *MCM3AP* variants can be roughly divided into two classes: missense variants in the Sac3 domain of GANP and compound heterozygous variants outside Sac3 (Supplementary material, Table S1). Homozygous missense variants in Sac3 have been described in several patients: p.Met762Thr in a Turkish family (2), p.Ser951Pro in an Iranian family (4) (IR cell line in this study), p.Arg878His in a Kurdish family (4), p.Glu915Lys in a Dutch family (1) and p.Leu870Ser in a Lebanese family (3). In addition, the Sac3 variant p.Ala867Asp has been found in combination with a nonsense variant p.Tyr899* in an Australian family (AUS cell line in this study) (2) (Fig. 1C). Furthermore, peripheral neuropathy and multiple sclerosis were associated with homozygous p.Ile954Thr variant (5). Amino acid conservation of the recessive Sac3 missense variants is shown in Supplementary material, Figure S3. To investigate *in silico* the location of these missense variants in the structure of the Sac3 domain, we exploited the high level of conservation between

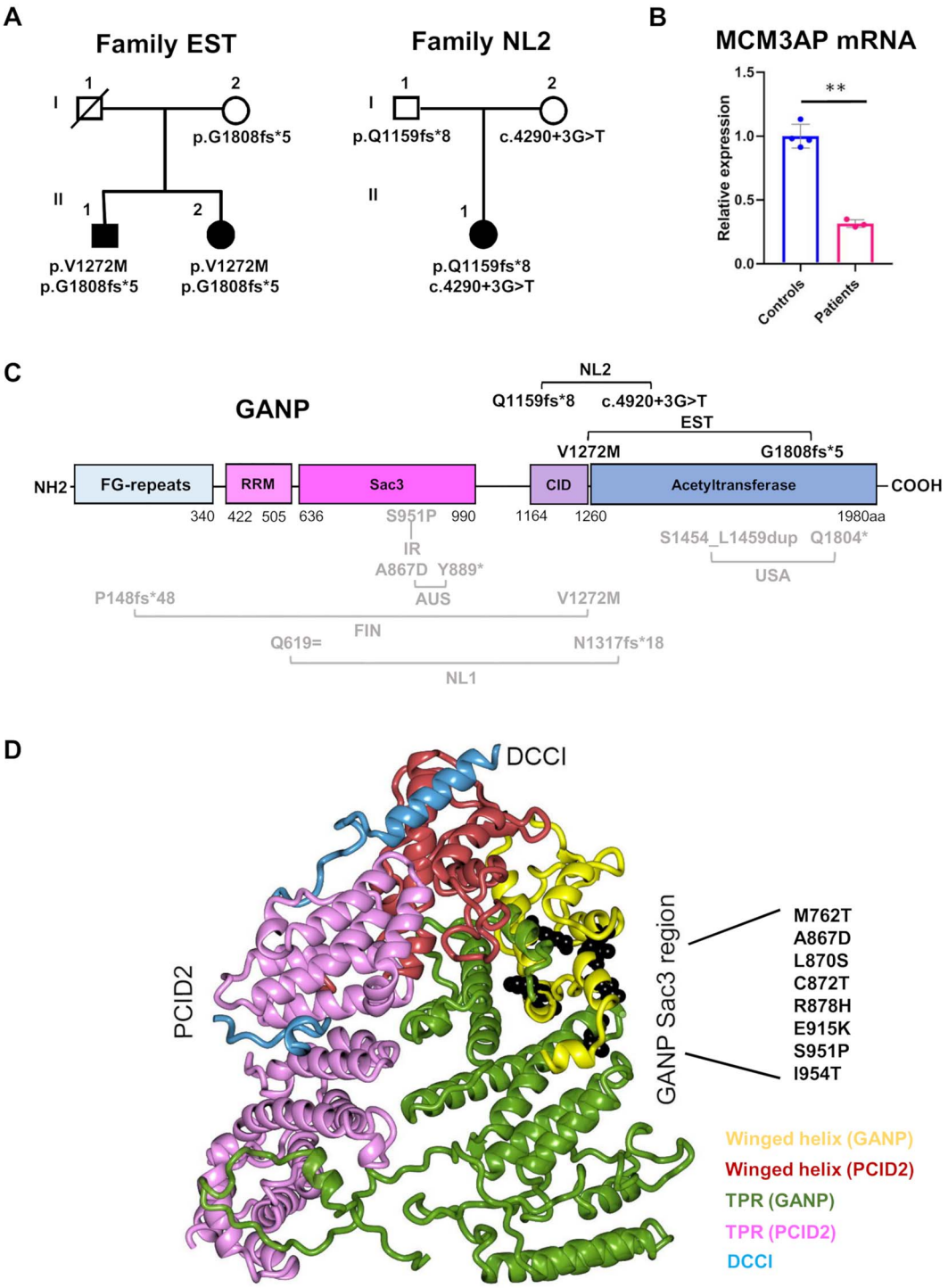

Figure 1. Novel variants in *MCM3AP* and structural modeling of Sac3 domain variants.
(**A**) Family pedigrees of EST and NL2. (**B**) qPCR quantification of *MCM3AP* mRNA levels in EST1, EST2 and NL2 patient and four control fibroblasts normalized to *GAPDH*. The mean of three technical replicates is shown for each cell line as dots. Bars show mean of control or patient cells. Error bars indicate SD. ** $P < 0.0001$. (**C**) Map of GANP variants investigated in this study. Variants in new families are shown in black and previously reported variants in grey. (**D**) Modelling of mutations in the Sac3 domain. The secondary structure of the GANP, PCID2 and DCCI chains is shown using different colors to identify specific components. The winged-helix regions of GANP and PCID2 are shown as yellow and red, respectively, their TPR (tetratricopeptide repeats) regions in green and magenta, respectively, and DCCI as light blue. All of the disease-associated missense variants (black) are located either in the GANP winged-helix domain (shown as yellow secondary structure) or in the GANP TPR region (shown as green) immediately adjacent to the winged helix domain. The residues mutated are primarily located in the interior of the structure and so are likely to generate local perturbations in the molecular conformation in this region. Although Ser951 is located on the surface, it lies within a helix, and because it is mutated to Pro, a residue that is incompatible with an α-helix, this mutation would also probably generate a local structural perturbation. It is likely that the local structural perturbations introduced by these mutations could influence the binding of RNAs to the GANP:PCID2:DCCI complex, either directly by changing the conformation of the GANP winged helix domain or indirectly by altering the interaction between GANP and PCID2.

the human GANP:PCID2:DCCI complex and the *Saccharomyces cerevisiae* Sac3:Thp1:Sem1 complex for which a crystal structure is available (27,28). The structures of both PCID2/Thp1 and GANP/Sac3 in this region have a PCI fold (29) based on a series of helical tetratricopeptide repeats (TPR) capped by a winged helix domain (Fig. 1D, Supplementary material, Movie S1). All but one of the Sac3 domain variants observed in the GANP patients are clustered in a region that lies in the winged helix domain close to its junction with the TPR region, whereas the remaining variant is in the TPR region immediately adjacent to the winged helix domain. Moreover, these variants are primarily in the interior of the protein and so probably generate local changes in structure. This winged helix region of Sac3, together with the adjacent PCID2, is associated with binding to RNA (28) and so local structural changes introduced by these variants would be expected to impair RNA binding either directly or indirectly by impairing the binding of PCID2.

Compound heterozygous *MCM3AP* variants outside the Sac3 domain lead to decreased GANP in nuclear envelope

To assess the consequences of *MCM3AP* variants on GANP protein, we studied skin fibroblasts from nine affected individuals from families FIN, AUS, USA, NL1, NL2, EST, IR and the unaffected mothers of NL1 (NLM) and IR (IRM) (Fig. 1C, Table 1). All other patients carried compound heterozygous variants outside the Sac3 domain, except the AUS patient, who had a heterozygous Sac3 missense variant together with a nonsense variant, and the IR patient, who carried the homozygous S951P variant in the Sac3 domain (Fig. 1C). We used a polyclonal antibody that recognizes the N-terminal region of GANP to perform immunostaining on these fibroblasts (Fig. 2A–C). Staining was done on three different occasions, upon receipt of the different patient cell lines (Fig. 2A–C). In summary, cells from affected individuals from families FIN, NL1, USA, EST and NL2 showed greatly reduced GANP staining at both the nuclear envelope and the nuclear interior when compared with control fibroblast lines (Fig. 2A and B). In contrast to all other patient lines, the AUS and IR cell lines carrying Sac3 missense variants had a GANP staining comparable with that of the controls (Fig. 2A and C). We quantified the GANP signal by manually counting the stained cells and grouping them depending on the intensity of staining at the nuclear envelope. GANP+ denoted a clear presence of GANP, GANP- denoted absence of GANP, and intermediate cells had faint GANP staining at the nuclear envelope (Fig. 2D). The quantification showed the highest number of GANP+ cells in controls (from 61% to 100%) and AUS (91%), as well as IR/IRM (from 95% to 100%) consistent with the visual observation (Fig. 2E). Unaffected heterozygous carrier NLM had mostly GANP+ cells (64%). Fibroblast lines of FIN, NL1, NL2 and EST had about 1% of GANP+ cells, whereas USA fibroblasts had none (0%). The ratio of GANP staining in counted cells was significantly different in all patients with variants outside the Sac3 domain (FIN1/2, EST1/2, NL1, NL2, USA) in comparison with controls ($P < 0.001$ for each patient), whereas patients with Sac3 variants did not differ from controls. Staining results of controls in the last two experiments are shown in Supplementary material, Figure S1. Western blot analysis of GANP, which was performed in some fibroblast lines, was consistent with the immunostaining results (Supplementary material, Fig. S2).

In conclusion, these results showed that compound heterozygous *MCM3AP* variants outside the Sac3 domain resulted in a decreased GANP level at the nuclear envelope. In contrast, patients with Sac3 domain variants had normal GANP expression and localization.

Nuclear envelope ultrastructure in the absence of GANP is normal

We used three fibroblast lines from affected (FIN1, FIN2 and AUS) and two from healthy individuals for ultrastructural assessment by transmission electron microscopy (TEM) (Fig. 3). The appearance of the nuclear envelope, nuclear pores and nucleoli and the general appearance of cells were normal in the GANP-depleted fibroblasts (FIN1/2) and in AUS cells carrying the Sac3 missense variant.

Affected individuals share a transcriptome signature in fibroblasts

Next, we assessed whether disease-associated GANP defects influenced gene expression in fibroblasts. mRNA from fibroblast lines FIN1/2, AUS, USA, NL1, NLM and two unrelated controls (one male and one female) were deep sequenced using paired-end sequencing using two technical replicates. We compared the overall gene expression profiles by hierarchical correlation analysis (Fig. 4A). To test if any gene expression alteration correlated with GANP deficiency, we performed differential gene expression (DE) analysis using DESeq2 (30) comparing patients with healthy controls. This analysis identified 51 downregulated and 110 upregulated genes (compared with control fibroblasts, FDR < 0.01) (Fig. 4B). *MCM3AP* mRNA levels were significantly reduced ($\log_2$FC = −0.81, FDR = 1.32^{-06}), as expected based on qPCR results.

To identify potential pathways affected by GANP deficiency, we used REACTOME, KEGG, DO and GO enrichment analysis (Fig. 4D). Although this study was done using patient fibroblasts, which are not the affected cell type in these patients, DO gene sets identified pathways related to brain developmental disorders and epilepsy, and REACTOME identified pathways related to neuronal processes. In particular, altered genes were related to neuronal adhesion (*NRXN3*), protein–protein interactions at synapses (*SLITRK1*, *NLGN1* and *APBA2*), potassium channels (*KCNA7*, *KCNMA1* and *KCNQ5*) and other neuronal and synaptic processes (*RIMS1, GRIA1, GABRB3, PTPRD* and *CACNB4*) (Fig. 4C).

GANP has been previously suggested to be selectively involved in the export of highly expressed and rapidly regulated genes, which tend to be shorter and have fewer introns than genes in general (9). Therefore, we examined the number of introns in the identified DE genes. Interestingly, we noticed that the distribution of exons was significantly different between up- and downregulated genes (Fig. 5A), by Wilcoxon rank sum test with continuity correction (Mann–Whitney *U* test W = 1772.5, *P*-value < 0.0002). For example, of the 51 downregulated DE genes, 15 were intronless, whereas all upregulated genes had introns (Fig. 5B). We validated the downregulation of three intronless genes *ATXN7L3B* ($P < 0.0001$), *TSPYL4* ($P < 0.0001$) and *EID1* (<0.0001) by qPCR, including eight patient lines in the analysis in comparison with control cells (Fig. 5C).

Taken together, these results suggest that GANP defects in patient fibroblasts affect the expression of a limited number of genes, and that intron content is a contributing factor.

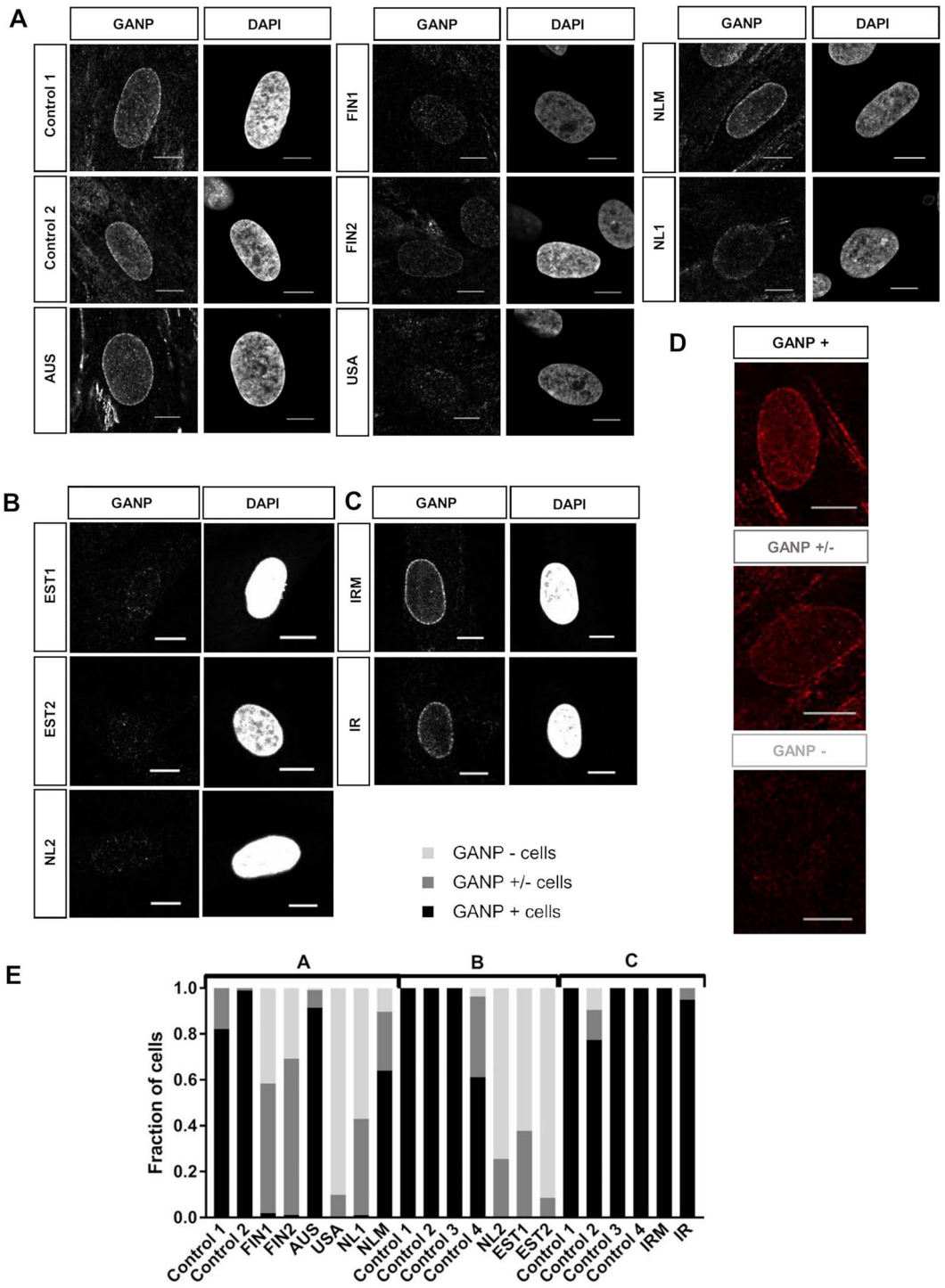

Figure 2. Variants outside Sac3 domain cause loss of GANP protein.
Immunocytochemistry with an N-terminal GANP antibody in patient fibroblasts. (**A**) Confocal images of patient group A (AUS, FIN1, FIN2, USA and NL1) and control fibroblasts stained with GANP antibody show normal localization to the nuclear envelope and interior in control and AUS cells and reduction in staining in other affected individuals' cell lines. NLM refers to cell line of mother of NL1. (**B**) Confocal images of patient group B (EST1, EST2 and NL2). (**C**) Confocal images of patient group C (IR). IRM refers to cell line of mother of IR. (**D**) Example of categorization of cells into GANP +, GANP intermediate or GANP –. (**E**) Bar graph shows quantification of GANP protein levels relative to GAPDH for A-C. Bar graph quantification of GANP staining in the nucleus and the nuclear envelope as fraction of cells from blindly counted cells for that sample. Cells were classified into GANP + (black), intermediate (–/+) (dark grey) and GANP—(light grey) cells based on staining in the nuclear envelope and interior. Quantification is based on technical triplicate samples of the same experiment $n^{average} = 112$ for samples A and technical duplicates $n^{average} = 136$ for samples B, $n^{average} = 90$ for duplicate samples in experiment C. Dapi stains nuclei. Scale bars: 10 µm. See Supplementary material figure S1A for control samples for experiments B and C.

Table 1. Human fibroblast lines used in this study

Affected individual designation (current study)	Affected individual designation (original study)	MCM3AP variants	Reference
FIN1	F:II-3	p.V1272M and p.P148Lfs*48	(2)
FIN2	F:II-4	p.V1272M and p.P148Lfs*48	(2)
AUS	A:II-2	p.A867D and p.Y889*	(2)
NL1	N:II-1	p.Q619= and p.N1317fs*18	(2)
NLM	N:I-2	Heterozygous p.N1317fs*18	(2)
USA	C-III.1	p.S1454_L1459dup and p.Q1804*	(4)
NL2	-	p.Q1159Rfs*8 and c.4290+3G>T	This study
EST1	-	p.V1272M and p.G1808Dfs*5	This study
EST2	-	p.V1272M and p.G1808Dfs*5	This study
IR	-	Homozygous p.S951P	(4)
IRM	-	Heterozygous p.S951P	(4)

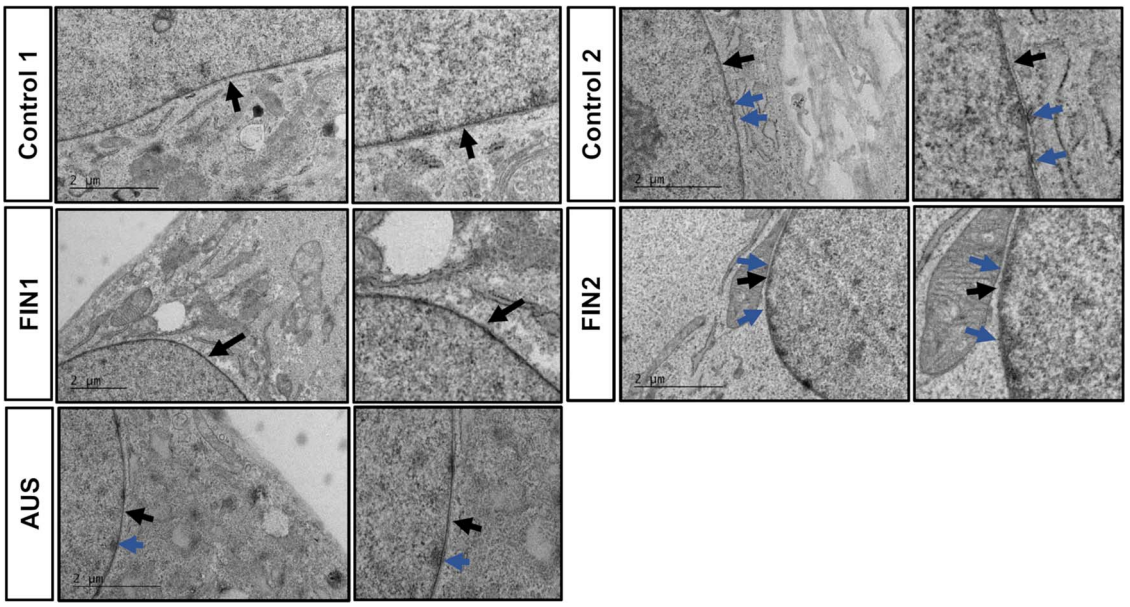

Figure 3. Ultrastructural assessment of the nuclear pore with electron microscopy.
Electron micrographs of patient FIN1/2, AUS and control fibroblasts. Blue arrows indicate nuclear pore complexes and black arrows the nuclear envelope. Scale bars: 2 µm.

ATXN7L3B mRNA expression but not localization is altered in patient cells

ATXN7L3B protein, encoded by one of the downregulated intronless genes (*ATXN7L3B*), interacts with ENY2 that functions as an adaptor protein in both the SAGA and TREX-2 complexes (31). Our qPCR analysis showed that *ATXN7L3B* mRNA levels were reduced by 39% on average in patient fibroblasts (Fig. 5C). Interestingly, in line with this result, we noticed that the GEO (GSE54481) dataset of GANP siRNA knockdown in human HCT116 cells also indicated downregulation of *ATXN7L3B* ($P = 1.69e-09$, adj.p.val = 8.43e-06, logFC = −0.935) (9).

To further study whether the *ATXN7L3B* mRNA localization was altered in patient cells, we used RNA *in situ* hybridization in fibroblasts from nine available patients and three controls with an *ATXN7L3B* specific probe. RNA quality of the samples was validated by using a probe targeting cyclophilin B (*PPIB*), and background staining was assessed by a probe targeting bacterial gene dapB. All samples expressed a high level of the *PPIB*, and no signal was detected for dapB. In control cells, *ATXN7L3B* probe showed a strong punctate signal across the proximal and distal regions of the cytoplasm, whereas in patient cells, the cytoplasmic signal was reduced (Fig. 6A). Quantitation showed that the average number of puncta per cell was significantly lower in patient cell lines compared with control cell lines ($P < 0.05$) (Fig. 6B). No signs of *ATXN7L3B* mRNA retention in the nucleus were noted in patient or control cells.

Although other SAGA or TREX-2 subunits were unaltered in RNASeq analysis, we additionally validated the expression levels of *ENY2*, *USP22* and *ATXN7L3* using qPCR on the full set of patient lines. In agreement with the RNASeq results, *ENY2* and *USP22* were not altered; however, *ATXN7L3* expression was increased by 27% on average in patient cells in comparison with controls ($P = 0.02$) (Supplementary material, Fig. S4).

Discussion

Disruptions in RNA metabolism are increasingly recognized as causes of a range of neurological phenotypes (32). Owing to the complexity of mRNA life cycle and the interdependence of its different steps, it is not surprising that even small changes in

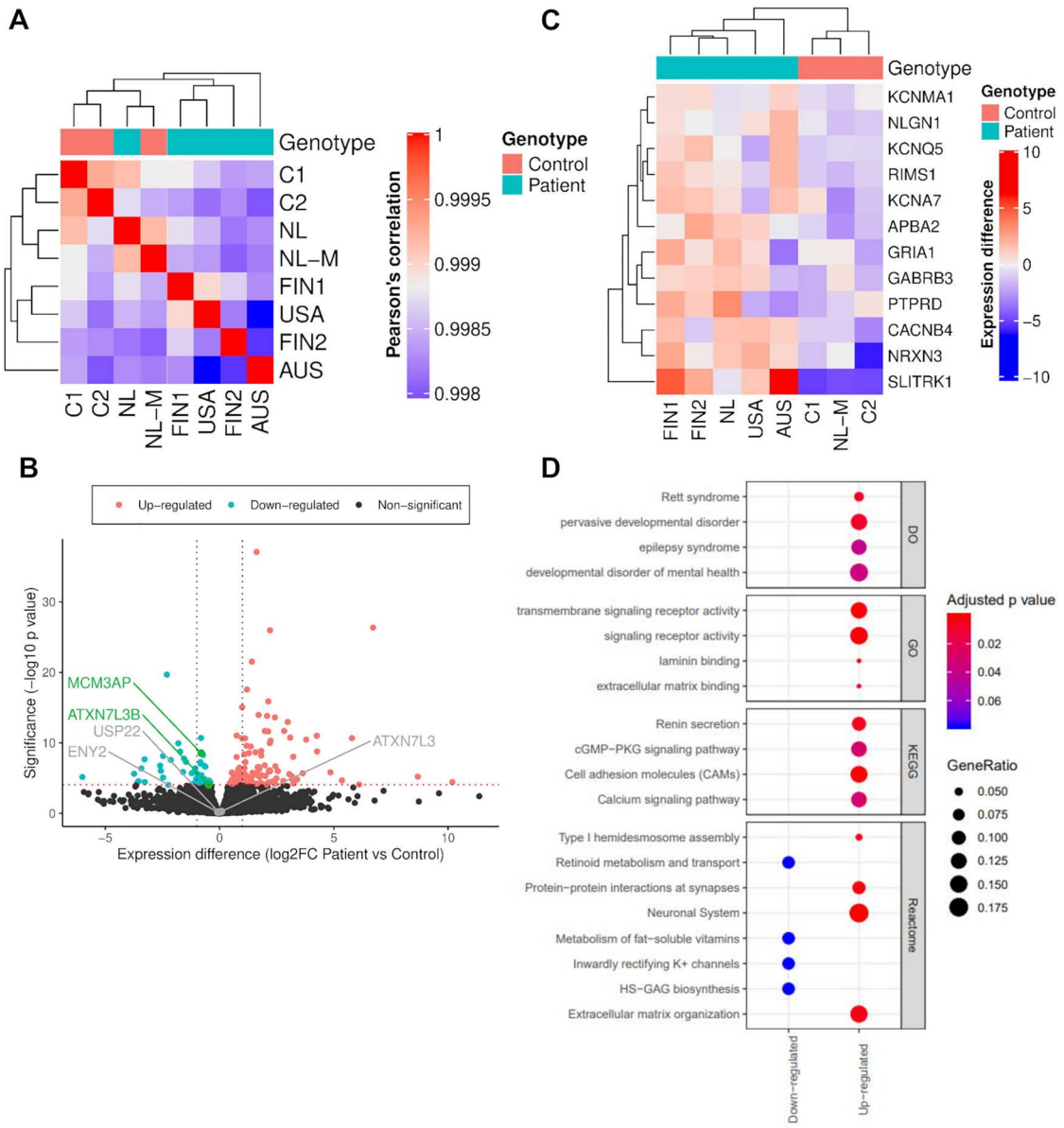

Figure 4. Differentially expressed genes in patient fibroblasts.
(**A**) Pearson hierarchical correlation of gene expression of affected individuals' and control fibroblasts. C indicates control. (**B**) Volcano plot showing expression differences and significance for genes. Of the TREX-2/SAGA complex associated genes *MCM3AP* and *ATXN7L3B* are indicated in green and non-significant *USP22*, *ENY2* and *ATXN7L3* in grey. (**C**) Heatmap of expression differences (to gene average) for the enriched genes in REACTOME category 'neuronal system'. See full list in Supplementary material Table S2. (**D**) Gene-set enrichment analysis. See full list in supplementary material Table S3.

the amount or activity of the machinery responsible may have widespread deleterious effects in neurons. Here we studied a protein recently associated with human neurological disease, GANP, which contributes to mRNA maintenance pathways, and to symptoms of peripheral and central nervous system disease.

Several separate studies have now identified biallelic *MCM3AP* mutations as causes of neuropathy with or without intellectual disability (1,2,3,4,5). The phenotype of the EST patients in this study was similar to that of the previously described Finnish patients who carried the same p.Val1272Met variant combined with a different truncating variant. All patients from these two families had early disease onset with upper motor neuron signs, and lost ambulation in adolescence or early adulthood. The Estonian patients had additionally progressive encephalopathy starting around ages 11–12 years. The patient NL2 had a severe motor phenotype with early central involvement. These findings increasingly indicate a role for GANP in central nervous system in addition to the peripheral nerves.

Because it forms the scaffold of the TREX-2 complex, GANP has a crucial role in mediating the export of mRNAs from nucleus to cytosol and most likely also influences the upstream steps of mRNA transcription and processing (7,9). However, GANP function and regulation in human cells is not fully understood. Our molecular characterization of patient fibroblasts with various GANP defects showed that compound heterozygous *MCM3AP* variants outside the mRNA binding Sac3 domain led to a dramatic decrease of the GANP protein at the nuclear envelope.

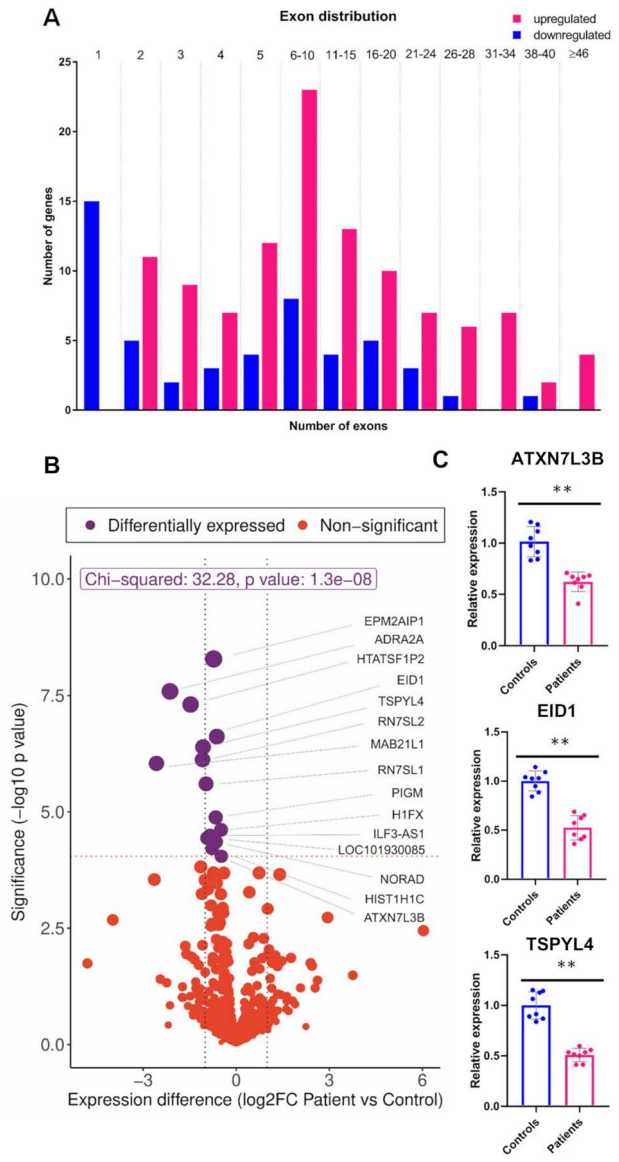

Figure 5. Downregulation of intronless genes in patient fibroblasts.
(**A**) Distribution of exons across differentially expressed genes. (**B**) Volcano plot indicating expression differences for up- and downregulated intronless genes in log2FC in patient versus control samples and significance as –log10 p-value. (**C**) qPCR validations of intronless genes *ATXN7L3B*, *EID1* and *TSPYL4* relative to *GAPDH*. Each dot represents average of three technical replicates of an affected (n = 8) or control individual (n = 4). ** $P < 0.0001$. Error bars show mean with SD.

In cells from these individuals, GANP was nearly undetectable by immunostaining, whereas fibroblasts from Sac3 variant carriers had normal GANP levels. Modeling of Sac3 disease variants in the *S. cerevisae* Sac3:Thp1:Sem1 complex suggested that the Sac3 variants associated with human disease may generate local structural changes that could in turn impair binding to RNA and so could provide a pathogenic mechanism in these cases when the levels of GANP protein remained near normal. Overall, our studies indicate that *MCM3AP*-associated mutations cause loss of GANP function. The unexpectedly strong depletion of GANP in the compound heterozygous patients explains why the associated phenotypes are more severe than those caused by Sac3 variants. Nevertheless, the phenotypes caused by both mutation types are quite similar, suggesting that the loss of mRNA binding function of GANP is a principal mechanism by which neuronal function is impaired.

Although skin fibroblasts are not the cell type affected in the *MCM3AP*-associated disease, detecting the severe GANP loss in the patient cells prompted us to study whether depletion of GANP has general effects on mRNA metabolism in fibroblasts. The neuronal disease could result from a molecular defect that is present in all cell types, but which only manifests in selected neurons due to profound downstream consequences of selected transcripts that are highly important in those neuron types. In line with GANP depletion having an effect, we found that fibroblasts from multiple affected individuals with different pathogenic variants shared similar transcriptional fingerprints. Higher proportions of differentially expressed genes were upregulated than downregulated, suggesting compensatory effects upon GANP depletion. Unexpectedly, many of the upregulated genes were involved in neuronal functions. This could suggest that GANP has a specific role in the expression of neuronal transcripts that result in their misregulation even in a non-neuronal cell type. Future studies using GANP-deficient human neurons derived from induced pluripotent stem cells are required to investigate the involvement of these neuronal transcripts further. Of note, potassium channels within the differentially expressed genes may be of interest, as hypokalemic periodic paralysis was associated with a heterozygous missense Sac3 variant in one family (6).

GANP has previously been found to have a role in selective mRNA export, affecting the transport of mRNAs differentially, with preference on genes that require less processing, that is, have less introns (9). We studied here the global effect of disease-associated GANP variants on steady-state mRNA levels, and not the effect on transport itself. We found no overall indication of GANP defects having a stronger effect on transcripts with fewer introns, but interestingly the distribution of intron content was different between significantly altered upregulated and downregulated genes, with downregulated genes having fewer introns. These findings suggest that GANP may have a role in the regulation of transcription and splicing processes in human cells in addition to mRNA export. In particular, we observed that many of the downregulated genes were intronless, whereas all upregulated genes had introns. Whereas it is known that intron-containing genes are surveyed by quality control mechanisms to ensure that unspliced mRNAs are not exported (17), the export mechanisms of intronless genes have not been investigated in great detail. TREX has been suggested to be involved in export of intronless mRNAs in HeLa cells (33), NXF1/TAP pathway in exporting intronless histone H2A mRNA (34) and nuclear pore basket protein TPR in export of intronless transcripts (35). *ATXN7L3B* was one of the downregulated intronless genes in patients' cells where RNA *in situ* hybridization showed a reduction in its overall amount but no change in mRNA localization and no evidence for nuclear retention. Therefore, there appears to be a clear feedback from defect in GANP to steady-state mRNA levels depending on whether the genes have introns to process. We speculate that in neurons, this effect on intronless genes may be rate-limiting for a set of key genes in development and maintenance of the nervous system. Furthermore, our results suggest that export of intronless mRNAs may require TREX-2 in addition to TREX.

In the RNASeq analysis of GANP-deficient patient fibroblasts, *ATXN7L3B* was the only altered transcript of the TREX-2, TREX or SAGA complex subunits. ATXN7L3B binds ENY2 (36), which is a key protein in both the TREX-2 and SAGA complexes (10,31,36,37). SAGA is a large chromatin modifying coactivator

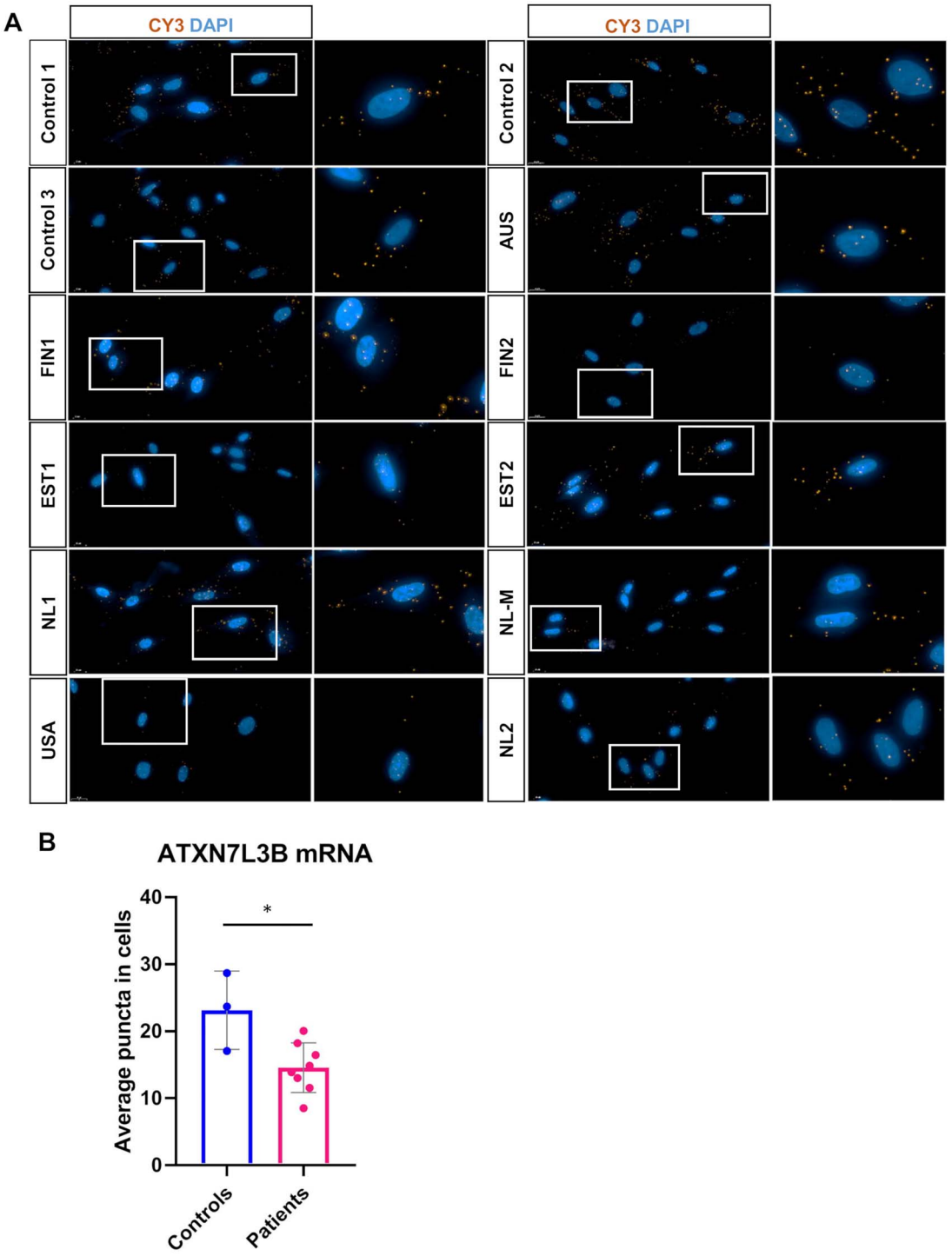

Figure 6. *ATXN7L3B* RNA *in situ* hybridization shows reduced signal in patient fibroblasts.
(**A**) Patient and control fibroblasts hybridized with *ATXN7L3B* target probe (Cy3) and dapi (blue). White square indicates focus area on the other image. Scale bars: 10 µm.
(**B**) Average of punctae number in cells/line is plotted in the bar graph. $P < 0.05$. Mean with SD plotted.

complex with modules for the acetylation (HAT) and deubiquitination (DUB) of histones as well as non-histone substrates. The DUB module removes monoubiquitin from K120 of human H2B (H2Bub1) in genes and has USP22 as the catalytic subunit and ATXN7L3 and ENY2 as adaptor proteins for activity and stability. A recent study found that ATXN7L3B, a cytoplasmic paralog of ATXN7L3, expressed from a pseudogene, competes with it for ENY2 binding, thus inhibiting the ability of SAGA complex to deubiquitinate histones (36). Whereas we found decreased *ATXN7L3B* expression in patient cells, qPCR analysis indicated increased *ATXN7L3*, supporting a connection between these two factors. Recent study has interestingly linked the GANP and SAGA connection to genome stability, by showing that a functional interaction between human TREX-2 and the

DUB activity of SAGA is needed to maintain the H2B/HB2Bub1 balance for efficient DNA repair and homologous recombination (38). These authors showed that, whereas cells lacking ENY2 had increased H2Bub1 levels due to decreased deubiquitination, GANP-depletion decreased global H2Bub1. Whether the human disease-associated GANP defects have an effect on SAGA function requires further studies. Interestingly, *ATXN7L3B* may well have relevance for the neurological phenotypes associated with GANP defects, as two studies have described aberrations affecting the chromosomal region for *ATXN7L3B* with links to neurological phenotypes (39,40). In conclusion, our results suggest that GANP is a regulator of *ATXN7L3B* expression.

In summary, we describe central nervous system involvement in new affected individuals with *MCM3AP* variants and provide a molecular explanation for the genotype/phenotype correlation through investigation of multiple cell lines from patients with GANP defects. Our results suggest a novel link between mRNA export and intron-dependent regulation of gene expression in human disease.

Materials and Methods

Fibroblast cell culture

Human fibroblast cultures were established from skin punch biopsies and cultured by standard methods. Briefly, cells were cultured in 37°C in 5% CO_2 and 95% O_2 in DMEM media (#BE12-614F, Lonza), 10% FBS, 1% penicillin–streptomycin, 1% L-glutamate and 0.1% uridine. Low-passage fibroblasts of affected individuals NL1 (female), USA (female), FIN1 (female), FIN2 (male), AUS (female), NL2 (female), EST1 (male) and EST2 (female), IR (female), and NLM referring to mother of NL1 and IRM to mother of IR and two to four healthy volunteer controls were used for the experiments (controls are named per experiment). The affected individuals or their guardians gave a voluntary informed consent to participate in the experiments.

Exome and genome sequencing

For EST family, whole genome sequencing and data processing were performed by the Genomics Platform at the Broad Institute of MIT and Harvard. PCR-free preparation of sample DNA (350 ng input at >2 ng/ul) is accomplished using Illumina HiSeq X Ten v2 chemistry. Libraries are sequenced to a mean target coverage of >30x. Genome sequencing data were processed through a pipeline based on Picard, using base quality score recalibration and local realignment at known indels. The BWA aligner was used for mapping reads to the human genome build 38. Single Nucleotide Variants (SNVs) and insertions/deletions (indels) are jointly called across all samples using Genome Analysis Toolkit (GATK) HaplotypeCaller package version 3.4. Default filters were applied to SNV and indel calls using the GATK Variant Quality Score Recalibration (VQSR) approach. Annotation was performed using Variant Effect Predictor (VEP). Lastly, the variant call set was uploaded to seqr for collaborative analysis between the Broad Center of Mendelian Genomics and investigator.

For NL2 family, exome sequencing was outsourced to GenomeScan (Leiden, The Netherlands). In brief, exomes were captured using the Agilent SureSelectXT Human all Exon v5 library kit (Agilent, Santa Clara, USA) accompanied by Illumina paired end sequencing on the Hiseq2500 platform (Illumina, San Diego, USA), generating 2×150 bp paired end reads with at least 80x median coverage. Data analysis was performed at the LDGA using the in-house sequence analysis pipeline 'Modular GATK-Based Variant Calling Pipeline' (MAGPIE) (LUMC Sequencing Analysis Support Core, LUMC) based on read alignment using Burrows-Wheeler Alignment (BWA-MEM) and variant calling using the Genome Analysis Toolkit (GATK). LOVDplus (Leiden Genome Technology Center, LUMC, Leiden) was used for interpretation of variants. Variants were classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines.

Fibroblast RNA sequencing and bioinformatics analysis

RNA was isolated from 90 to 100% confluent 10-cm plates, grown in culture for 4 days. Technical duplicates were sequenced and analyzed for all cell lines, which were passaged at the same time. All the fibroblast lines were in early passage stages and grew at approximately equal pace. Cells were washed with PBS once, scraped and centrifuged 1000 rcf for 5 min. Pellets were immediately processed for RNA isolation with Nucleospin RNA kit (#740955.50, Macherey-Nagel) according to the manufacturer's instructions including rDNAse treatment. RNA concentrations were measured with Agilent TapeStation and Nanodrop, RNA integrity was checked on a gel with Agilent TapeStation and RIN were ≥ 9.6. RNA measurements and mRNA sequencing were done at Biomedicum Functional Genomics Unit with paired-end sequencing with Poly A binding beads and NEBNext Ultra Directional RNA Library Prep and sequenced with Illumina NextSeq High Output 2 x 75 bp. Target number of reads was 45–50 M, which was ensured by sequencing with Next Seq Mid Output flow cell.

The raw data were demultiplexed with adapter removal using bcl2fastq2 (v2.20.0.422; Illumina). FASTQ-files were filtered with Trimmomatic (v.0.36; (41)) excluding non-paired, short (<35 bp) and low quality (phread score <20) reads. The filtered reads were mapped to the human reference genome (GRCh38) with STAR (v. 2.5.3; (42)). Expression analysis was carried out in R (R Development Core Team & Core, 2018). Read counts for genes were extracted with Rsubreads (v.1.32.4; (43)) excluding duplicates, multimapping reads, chimeric fragments and reads with low mapping quality (<10). Differential gene expression was analyzed using DESeq2 (v.1.22.2; (30)), comparing disease status (control versus patient). The technical replicates were combined, after confirming that the expression profiles were nearly identical. Genes with low read depth (<50 reads in total) were excluded from the analysis. The results from the expression analysis along with the raw sequence data were deposited to GEO, under accession GSE135098.

cDNA synthesis and RT-PCR

Fibroblast total RNA was isolated with Nucleospin RNA kit (#740955.50, Macherey-Nagel). cDNA was synthesized according to the manufacturer's instructions with Maxima First Strand cDNA Synthesis Kit (#K1671, Thermo Fisher Scientific).

qRT-PCR

Quantitative Real Time PCR for quantification of *MCM3AP* mRNA amount, as well as RNA sequencing validations was done with cDNA isolated as described above. qPCR was run on a CFX96 Touch Real-Time PCR detection system (Bio-Rad) with SYBR Green qPCR Kit (#F415S, Thermo Fisher Scientific) on a 96-well plate with gene-specific primers 10 μM and H_20. The cycling protocol was 95°C for 7 min, then 40 cycles of 95°C for 10 s and

60°C for 30 s. The primers were tested to fit on a standard curve. Normalization was done with GAPDH as a housekeeping gene. Technical triplicates were used for all studies. 12.5 ng of cDNA was used.

Immunocytochemistry, fluorescence microscopy and confocal microscopy

Fibroblasts were grown in standard cell culture conditions as described above on cover glasses in 6- or 24-well plates for 4 days, and media was changed every other day. The confluent cells were washed three times with PBS, fixed with 4% paraformaldehyde for 15 min and washed three times with PBS. Cells were then permeabilized in 0.2% Triton X-100 in PBS for 15 min in room temperature. Coverslips were washed three times in PBS-T (0.1% Tween-20) followed by blocking in 5% BSA/PBS-T (#001-000-162, Jackson Immuno Research) for 2 h in room temperature on a shaker. Primary antibody in 5% BSA/PBS-T was incubated overnight at 4°C. The following day, coverslips were washed with PBS-T for 15 min three times, followed by incubation in secondary antibody in 5% BSA/PBST for 1 h on a shaker. Finally, the cells were washed three times for 10 min and plated on glass with Vectashield (#H-1200, Vector Laboratories).

Fibroblasts with GANP staining were quantified by manually counting under a fluorescent microscope (Zeiss). The number of GANP positive and negative cells was counted in a blinded manner from triplicate samples, from dapi positive nuclei, $n^{mean} = 112$ (first experiment) $n^{mean} = 136$ (second experiment) $n^{mean} = 90$ duplicate samples (third experiment). The groups were GANP positive, GANP negative and moderately GANP positive nuclei. The fibroblasts were imaged with Zeiss LSM880 IndiMo Axio Observer laser scanning confocal microscope with ZEN black software with DyLight 594 (GANP) imaged with laser 561 and Dapi imaged with laser 405 using the same exposure time for all samples. Imaging objective was Plan-Apochromat 63x/1.4 Oil Dic M27, and images were acquired at 8-bit depth and 512 x 512 (experiments B and C) or 1380 x 1380 (experiment A) frame size. The settings for the 561 laser (excitation 561, emission 659) were: 1.4% laser power, binning 1 x 1. The results were visualized with Zen Blue, Image J and Graph Pad Prism Software.

RNA *in situ* hybridization

RNA *in situ* hybridization was performed on fixed adherent fibroblast cells grown on four-chamber slides using RNAscope Multiplex Fluorescent Reagent Kit Version 2 (#323100, Advanced Cell Diagnostics) for target detection according to instructions. Duplicates were analyzed for target probe and single wells for negative and positive controls per sample. In short, the cells were fixed in 10% neutral buffered formalin and dehydrated in increasing concentrations of ethanol (50%, 70% and absolute ethanol). The cells were stored in absolute ethanol in −20°C until performing the RNA *in situ* hybridization. Next, the cells were rehydrated in decreasing concentrations of ethanol (70% and 50%) and washed in PBS. The probes *PPIB* (positive control, #313901), dapB (negative control, #310043) and *ATXN7L3B* (target probe, #804431) were hybridized for 2 h at 40°C followed by signal amplification and developing the HRP channel with TSA Plus fluorophore Cyanine 3 (1:1500 dilution) (NEL744001KT, Perkin Elmer) according to the manual. The cells were counterstained with DAPI and mounted with ProLong Gold Antifade Mountant (P36930, Invitrogen). The samples were scanned with 3DHISTECH Pannoramic 250 FLASH II digital slide scanner using 40x objective and extended focus option with seven focus levels. Punctae number was counted on Pannoramic viewer per cell and averaged per sample (average 24 cells per sample). Images were generated using 3DHISTECH Pannoramic 250 FLASH II digital slide scanner at Genome Biology Unit supported by HiLIFE.

Western blotting

Fibroblasts for protein analysis were grown on 10-cm petri dishes until confluent. The cells were isolated with 1X RIPA (#9806, Cell Signaling) for 5 min on ice, scraped and centrifuged at 14 000*g* for 10 min at 4°C. Protein concentrations were measured with Bradford (#500-0006, Bio-Rad) from triplicate samples and with a standard curve. For SDS-PAGE analysis of affected and control fibroblasts, 30 µg of protein was loaded per well on a 10% TGX 10-well 30 µL/well gel (456-1033, Bio-Rad) for 1 h. at 150 V. Molecular size of bands was detected with protein standard (#161-0374, Bio-Rad) and proteins loaded together with 4x Laemmli loading buffer (#161-0747, Bio-Rad) in 2-mercaptoethanol (#1610710, Bio-Rad). Gel was transferred into a nitrocellulose membrane (#1704158, Bio-Rad) and transferred using High Molecular Weight protocol in Trans-Blot Turbo (Bio-Rad). The blot was washed and blocked in 5% milk/TBST for 1 h in RT. For immunoblotting analysis, GANP antibody (1:1000) (#113295, Abcam) was incubated for 48–72 h in 4°C in 2.5% milk/BSA in TBS-T. The blot was washed three times for 5 min in TBS-T. Secondary antibody α-rabbit (1:5000) in 1% BSA/TBS-T was incubated for 1 h at RT on a shaker and washed for three times a 10 min in TBS-T. The blot was visualized with ECL reaction with Pierce ECL Western Blotting Substrate (#32106, Thermo Fisher). Loading control GAPDH (#14C10) was incubated overnight in 4°C and processed otherwise in the same way. The images were obtained with a Chemidoc imager (Bio Rad) and quantified on Image Lab software (Bio-Rad). The quantification was done using GAPDH to normalize protein levels of GANP. Results were visualized on Graph Pad Prism Software.

Antibodies

α-rabbit N-terminal GANP antibody for immunocytochemistry was (#198173, Abcam), used at a ratio of 1:200. DyLight or Alexa 594 goat-anti-rabbit secondary (#35560, Thermo Fisher; Alexa fluor) was used to detect GANP at 1:1000. N-terminal antibody used for western blotting of α-rabbit GANP (#113295, Abcam) was used at a ratio of 1:1000. Secondary antibody to detect GANP in immunoblotting was goat-anti-rabbit (#111-035-144, Jackson Immuno Research), at a ratio of 1:5000.

Molecular modeling

The structure of GANP in the Sac3 region was based on the crystal structure of this region in the *S. cerevisiae* TREX-2 complex (27,28), PDB accession codes 3T5V and 5UBP. This modeling was justified because of the high sequence homology between the human and yeast proteins in this region. Images and movies were generated using the CCP4MG software package.

Transmission electron microscopy

Fibroblast cultures on glass slides were fixed with 2% PFA + 2% glutaraldehyde (Sigma) for 30 min at 4°C, washed with PBS and finally with 2% glutaraldehyde O/N at 4°C. Fixed cells were prepared according to standard protocols for transmission electron microscopy at the Electron microscopy Unit of the Institute of

Biotechnology. Images were captured with the Jeol 1400 TEM microscope at 80 000 V.

Statistics

Unpaired parametric two-tailed *t*-test was used for qRT-PCR and RNAScope experiments in which two biological groups were compared (affected individuals versus controls). Significance was measured by $P < 0.05$. Statistical tests were performed in Graph Pad Prism 7.04. Fisher's exact test was used for analysis of immunocytochemical staining data. For intron distribution testing Wilcoxon rank sum test with continuity correction (Mann–Whitney *U* test) was used. For analysis of GEO siRNA data, analysis was run in GEO2R browser comparing GANP KD and control samples, and statistics were reported from the results table. All measured data points were included in the analysis.

Data availability

RNA-Seq data: Gene Expression Omnibus GSE135098 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135098).

Funding

This work was supported by Doctoral Programme Brain & Mind (RW), University of Helsinki (EY, HT), Academy of Finland (EY, HT), Instrumentarium Science Foundation (RW), HUS Helsinki University Hospital (EY), Emil Aaltonen Foundation (EY), Sigrid Juselius Foundation (HT), Estonian Research Council grant PRG471. The Broad Center for Mendelian Genomics (UM1 HG008900) is funded by the National Human Genome Research Institute with supplemental funding provided by the National Heart, Lung, and Blood Institute under the Trans-Omics for Precision Medicine (TOPMed) program and the National Eye Institute. MHW is supported by T32GM007748. MS holds a Leverhulme Emeritus Fellowship (EM-2016-062) and is supported by MRC grants MC U105178939 and MC UP 1201/6. Work in the laboratory of CGB is supported by intramural funds of National Institute of Neurological Disorders and Stroke/National Institutes of Health.

Author Contributions

RW, EY and HT conceived the ideas and planned the experiments with the help of MW and JS. RW developed and modified the methodologies used. JK did the bioinformatics analysis and implemented the codes. RW completed and validated the experiments and did the formal analyses. NA performed RNAScope experiments, which OC supervised. MS performed molecular modeling experiments. FZ and RW analyzed electron microscopy data. KÕ, KR, MHW, EKB, MJH, MMR, ZS, MW, IC, MJB, DB-G, SD, SW, RZ, CGB, RM, EOC and HH identified affected individuals and disease variants, did clinical analysis and provided the samples. RW, JK and MS prepared the figures and visualization of the data. RW wrote the initial manuscript and MS, EY and HT reviewed and edited the draft. All authors reviewed the final manuscript.

Conflicts of Interest Statement

DB-G is a steering committee member for the Gene Therapy Network (sponsored by Avexis) and received a Career Development Award (American Society for Gene and Cell Therapy).

Acknowledgements

We thank the affected individuals and their families for participating in this study. Riitta Lehtinen and Jana Pennonen are acknowledged for technical assistance and Laura Mäenpää for initial RNASeq analysis. We acknowledge Biomedicum Functional Genomics Unit, Biomedicum Imaging Unit and Electron Microscopy Unit of the Institute of Biotechnology, University of Helsinki, and Biocenter Finland.

Supplementary Material

Supplementary material is available at *HMG* online

References

1. Schuurs-Hoeijmakers J.H., Vulto-van Silfhout A.T., Vissers L.E., van de V.,II, van Bon B.W., de Ligt J., Gilissen C., Hehir-Kwa J.Y., Neveling K., del Rosario M. *et al.* (2013) Identification of pathogenic gene variants in small families with intellectually disabled siblings by exome sequencing. *J. Med. Genet.*, **50**, 802–811.
2. Ylikallio, E., Woldegebriel, R., Tumiati, M., Isohanni, P., Ryan, M.M., Stark, Z., Walsh, M., Sawyer, S.L., Bell, K.M., Oshlack, A. *et al.* (2017) MCM3AP in recessive charcot-marie-tooth neuropathy and mild intellectual disability. *Brain*, **140**, 2093–2103.
3. Kennerson, M.L., Corbett, A.C., Ellis, M., Perez-Siles, G. and Nicholson, G.A. (2018) A novel MCM3AP mutation in a Lebanese family with recessive charcot-marie-tooth neuropathy. *Brain*, **141**, e66.
4. Karakaya, M., Mazaheri, N., Polat, I., Bharucha-Goebel, D., Donkervoort, S., Maroofian, R., Shariati, G., Hoelker, I., Monaghan, K., Winchester, S. *et al.* (2017) Biallelic MCM3AP mutations cause charcot-marie-tooth neuropathy with variable clinical presentation. *Brain*, **140**, e65.
5. Sedghi, M., Moslemi, A., Cabrera-Serrano, M., Ansari, B., Ghasemi, M., Baktashian, M., Fattahpour, A. and Tajsharghi, H. (2019) Recessive charcot-marie-tooth and multiple sclerosis associated with a variant in *MCM3AP*. *Brain Commun.*, **1**, fcz011.
6. Gustavsson, E., Follett, J., Farrer, M. and Aasly, J. (2019) Family with primary periodic paralysis and a mutation in MCM3AP, a gene implicated in mRNA transport. *Muscle Nerve*, **60**, 311–314.
7. Wickramasinghe, V.O., McMurtrie, P.I., Mills, A.D., Takei, Y., Penrhyn-Lowe, S., Amagase, Y., Main, S., Marr, J., Stewart, M. and Laskey, R.A. (2010) mRNA export from mammalian cell nuclei is dependent on GANP. *Curr. Biol.*, **20**, 25–31.
8. Abe, E., Kuwahara, K., Yoshida, M., Suzuki, M., Terasaki, H., Matsuo, Y., Takahashi, E.I. and Sakaguchi, N. (2000) Structure, expression, and chromosomal localization of the human gene encoding a germinal center-associated nuclear protein (GANP) that associates with MCM3 involved in the initiation of DNA replication. *Gene*, **255**, 219–227.
9. Wickramasinghe, V.O., Andrews, R., Ellis, P., Langford, C., Gurdon, J.B., Stewart, M., Venkitaraman, A.R. and Laskey, R.A. (2014) Selective nuclear export of specific classes of mRNA from mammalian nuclei is promoted by GANP. *Nucleic Acids Res.*, **42**, 5059–5071.
10. Jani, D., Lutz, S., Hurt, E., Laskey, R.A., Stewart, M. and Wickramasinghe, V.O. (2012) Functional and structural characterization of the mammalian TREX-2 complex that links

transcription with nuclear messenger RNA export. *Nucleic Acids Res.*, **40**, 4562–4573.
11. Jones, A.L., Quimby, B.B., Hood, J.K., Ferrigno, P., Keshava, P.H., Silver, P.A. and Corbett, A.H. (2000) SAC3 may link nuclear protein export to cell cycle progression. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 3224–3229.
12. Jani, D., Lutz, S., Marshall, N.J., Fischer, T., Kohler, A., Ellisdon, A.M., Hurt, E. and Stewart, M. (2009) Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export. *Mol. Cell*, **33**, 727–737.
13. Kopytova, D.V., Orlova, A.V., Krasnov, A.N., Gurskiy, D.Y., Nikolenko, J.V., Nabirochkina, E.N., Shidlovskii, Y.V. and Georgieva, S.G. (2010) Multifunctional factor ENY2 is associated with the THO complex and promotes its recruitment onto nascent mRNA. *Genes Dev.*, **24**, 86–96.
14. Gallardo, M., Luna, R., Erdjument-Bromage, H., Tempst, P. and Aguilera, A. (2003) Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J. Biol. Chem.*, **278**, 24225–24232.
15. Takei, Y., Swietlik, M., Tanoue, A., Tsujimoto, G., Kouzarides, T. and Laskey, R. (2001) MCM3AP, a novel acetyltransferase that acetylates replication protein MCM3. *EMBO Rep.*, **2**, 119–123.
16. Takei, Y., Assenberg, M., Tsujimoto, G. and Laskey, R. (2002) The MCM3 acetylase MCM3AP inhibits initiation, but not elongation, of DNA replication via interaction with MCM3. *J. Biol. Chem.*, **277**, 43121–43125.
17. Bonnet, A. and Palancade, B. (2015) Intron or no intron: a matter for nuclear pore complexes. *Nucleus*, **6**, 455–461.
18. Carey, K.T. and Wickramasinghe, V.O. (2018) Regulatory potential of the RNA processing machinery: implications for human disease. *Trends Genet.*, **34**, 279–290.
19. Helmlinger, D. and Tora, L. (2017) Sharing the SAGA. *Trends Biochem. Sci.*, **42**, 850–861.
20. Umlauf, D., Bonnet, J., Waharte, F., Fournier, M., Stierle, M., Fischer, B., Brino, L., Devys, D. and Tora, L. (2013) The human TREX-2 complex is stably associated with the nuclear pore basket. *J. Cell. Sci.*, **126**, 2656–2667.
21. Lang, G., Bonnet, J., Umlauf, D., Karmodiya, K., Koffler, J., Stierle, M., Devys, D. and Tora, L. (2011) The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. *Mol. Cell. Biol.*, **31**, 3734–3744.
22. Kurshakova, M.M., Krasnov, A.N., Kopytova, D.V., Shidlovskii, Y.V., Nikolenko, J.V., Nabirochkina, E.N., Spehner, D., Schultz, P., Tora, L. and Georgieva, S.G. (2007) SAGA and a novel drosophila export complex anchor efficient transcription and mRNA export to NPC. *EMBO J.*, **26**, 4956–4965.
23. Wu, J., Bao, A., Chatterjee, K., Wan, Y. and Hopper, A.K. (2015) Genome-wide screen uncovers novel pathways for tRNA processing and nuclear-cytoplasmic dynamics. *Genes Dev.*, **29**, 2633–2644.
24. Guria, A., Tran, D.D., Ramachandran, S., Koch, A., El Bounkari, O., Dutta, P., Hauser, H. and Tamura, T. (2011) Identification of mRNAs that are spliced but not exported to the cytoplasm in the absence of THOC5 in mouse embryo fibroblasts. *RNA*, **17**, 1048–1056.
25. Herold, A., Teixeira, L. and Izaurralde, E. (2003) Genome-wide analysis of nuclear mRNA export pathways in Drosophila. *EMBO J.*, **22**, 2472–2483.
26. Ylikallio, E., Woldegebriel, R. and Tyynismaa, H. (2018) Reply: a novel MCM3AP mutation in a Lebanese family with recessive charcot-marie-tooth neuropathy. *Brain*, **141**, e67.
27. Gordon, J.M.B., Aibara, S. and Stewart, M. (2017) Structure of the Sac3 RNA-binding M-region in the *Saccharomyces cerevisiae* TREX-2 complex. *Nucleic Acids Res.*, **45**, 5577–5585.
28. Ellisdon, A.M., Dimitrova, L., Hurt, E. and Stewart, M. (2012) Structural basis for the assembly and nucleic acid binding of the TREX-2 transcription-export complex. *Nat. Struct. Mol. Biol.*, **19**, 328–336.
29. Ellisdon, A.M. and Stewart, M. (2012) Structural biology of the PCI-protein fold. *Bioarchitecture*, **2**, 118–123.
30. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550. doi: 550-014-0550-8.
31. Rodriguez-Navarro, S., Fischer, T., Luo, M.J., Antunez, O., Brettschneider, S., Lechner, J., Perez-Ortin, J.E., Reed, R. and Hurt, E. (2004) Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell*, **116**, 75–86.
32. Nussbacher, J.K., Tabet, R., Yeo, G.W. and Lagier-Tourenne, C. (2019) Disruption of RNA metabolism in neurological diseases and emerging therapeutic interventions. *Neuron*, **102**, 294–320.
33. Lei, H., Dias, A.P. and Reed, R. (2011) Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 17985–17990.
34. Huang, Y., Gattoni, R., Stevenin, J. and Steitz, J.A. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol. Cell*, **11**, 837–843.
35. Lee, E.S., Wolf, E.J., Smith, H.W., Emili, A. and Palazzo, A.F. (2019) TPR is required for the nuclear export of mRNAs and lncRNAs from intronless and intron-poor genes. *BioRxiv* (preprint).
36. Li, W., Atanassov, B.S., Lan, X., Mohan, R.D., Swanson, S.K., Farria, A.T., Florens, L., Washburn, M.P., Workman, J.L. and Dent, S.Y.R. (2016) Cytoplasmic ATXN7L3B interferes with nuclear functions of the SAGA deubiquitinase module. *Mol. Cell. Biol.*, **36**, 2855–2866.
37. Koutelou, E., Hirsch, C.L. and Dent, S.Y. (2010) Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol.*, **22**, 374–382.
38. Evangelista, F.M., Maglott-Roth, A., Stierle, M., Brino, L., Soutoglou, E. and Tora, L. (2018) Transcription and mRNA export machineries SAGA and TREX-2 maintain monoubiquitinated H2B balance required for DNA repair. *J. Cell Biol.*, **217**, 3382–3397.
39. Luukkonen, T.M., Mehrjouy, M.M., Poyhonen, M., Anttonen, A.K., Lahermo, P., Ellonen, P., Paulin, L., Tommerup, N., Palotie, A. and Varilo, T. (2018) Breakpoint mapping and haplotype analysis of translocation t(1;12)(q43;q21.1) in two apparently independent families with vascular phenotypes. *Mol. Genet. Genomic Med.*, **6**, 56–68.
40. Rajakulendran, S., Roberts, J., Koltzenburg, M., Hanna, M.G. and Stewart, H. (2013) Deletion of chromosome 12q21 affecting KCNC2 and ATXN7L3B in a family with neurodevelopmental delay and ataxia. *J. Neurol. Neurosurg. Psychiatry.*, **84**, 1255–1257.
41. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
42. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
43. Liao, Y., Smyth, G.K. and Shi, W. (2013) The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.*, **41**, e108.